会話音声の音響処理部と言語処理部の検討 (時系列パターンの認識システムの研究)

AUTHOR(S):

好田, 正紀; 中津, 良平; 鹿野, 清宏

数理解析研究所講究録
229巻 1975年 166-181

会話音声の音響処理部と言語処理部の検討

電電公社 武蔵野通研　好田 正紀
中津 良平
鹿野 清宏

1. まえがき

音声パターンは，時系列パターンの典型的な例である．そこで，時系列パターン認識システムとして，我々が現在検討を進めている音声による質問応答システムをとり上げる．

音声による質問応答システムの概要に関して，はじめに，入力音声の音響処理部について述べる．次に，言語処理部における単語認識のための関連づけと構文解析について述べる．最後に，このシステムを時系列パターン認識システムとしてみると，特徴パラメータレベル，音韻レベル，単語レベルの3種類の時系列パターンを扱っていることを述べ，各レベルの時系列パターンの性質や処理方法などを比較検討する．

2. 音声による質問応答システムの概要

音声認識の研究は，音響レベルの処理を主とする単語音声認識から，言語レベルの処理をも行なう会話音声理解へと研

究の対象が広がっており[1][2]，我々は，音声応答と組み合わせて，図1のような音声理解応答システムを考えている．

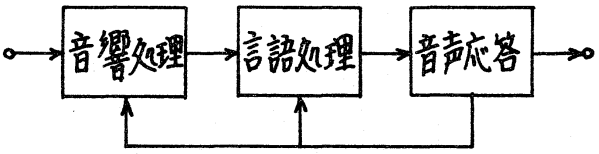


図1　音声理解応答システム

音声理解応答システムを構成して問題点を具体的に検討するために，質問応答システムの一例として，ここでは，荷物の座席予約をとり上げることとする[3]．

予約項目は，荷物の発駅，着駅，発時刻，荷物名，等，枚数の6項目とする．予約は表1に示されるような質問応答形式で行ない，この例にみられるように，問い合わせた事柄の一部しか入力されなかったり，あるい

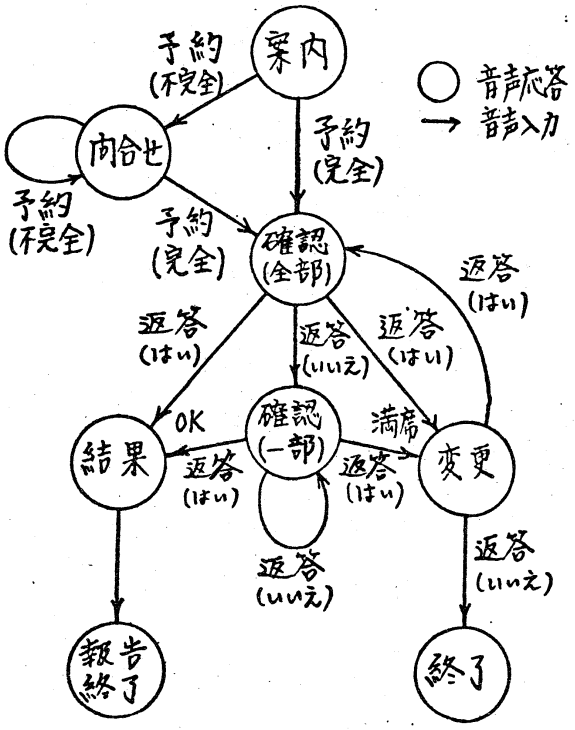


図2　座席予約の会話状態遷移図

表1　荷物座席予約における質問応答の例

音声応答	音声入力
・御希望をどうぞおっしゃって下さい	・A駅から，B駅まで，下さい．
・何時何分発で何枚ですか	・C時D分発で，E号，です．
・何枚ですか	・F枚，です．
・あなたの予約はE号A駅C時D分発B駅までの普通券をG枚ですね	
・F枚ですね	・いいえ，F枚，です．
	・はい，そうです．

は，別の事柄が入力されたりする場合も許すこととする．入力文は予約項目の節ごとに区切って発音する．一つの節は１～１０個の単語からなっている．会話のやりとりは図２に示されるような状態遷移図で表わすことができる．入力の語彙は，駅名に１３個の単語を用いることにすると，その他の名詞，助詞，副詞，動詞などを加えて合計６０語余りになる．

本システムは，音響処理，言語処理，音声応答の３つの部分からなる．それらの概要について次に述べる．

2.1 音響処理(4)(5)

音響レベルの処理を行なうとき，処理のための音声単位としては，単語よりも小さいＶＣＶ（母音－子音－母音）音節とする．ＶＣＶ音節は，

（1）単語よりも小さいため語彙の追加，変更が容易である，

（2）前後の母音ではさまれた子音が調音結合の影響を最も受けやすい，

ことを考慮して選んだ音声単位であって，会話音声の処理に適していると考えている．

音響処理部のブロック図を図３に示す．入力を文献（４）の方法

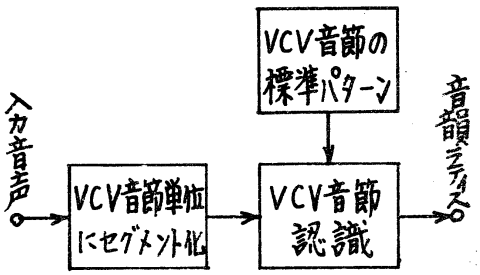


図３　音声理解応答システムの音響処理部

でVCV音節ごとにセグメント化する．VCV音節セグメントの認識は，このセグメントを15ミリ秒ごとに相関分析して得られる自己相関係数の時系列パターンと，VCV音節の標準パターンとして貯えられている最尤スペクトルパラメータの時系列パターンとを，図4に示されるように，DPによる時間軸の非線形正規化マッチングすることによって行なわれる．

この結果，入力音声はVCV音節の系列を経て音韻系列に変換される．しかしながら，こうして得られる音韻系列には種々の誤りが含まれており，入力音声に関する多くの情報が失われていることが多い．そこで，音響処理の結果を，ここでは，音韻系列における音韻のあいまいさとセグメンテーションのあいまいさの両方を許した図5のような音韻のセグメントラティスの形式で表現することとした．音韻のあいまい

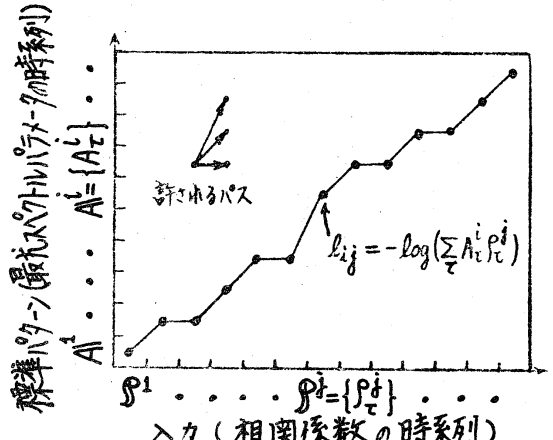


図4 VCV音節の認識における特徴パラメータ系列パターンのDPによるマッチング

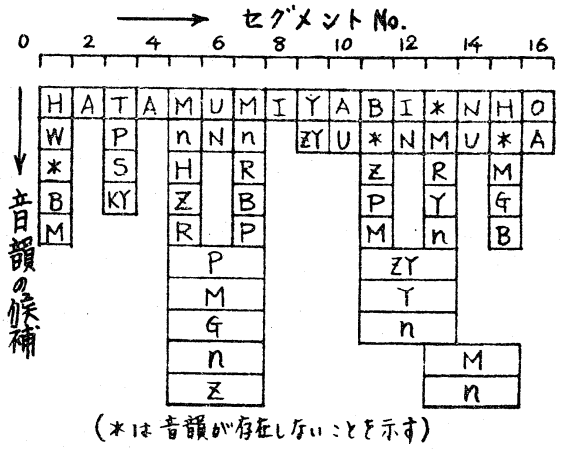


図5 音韻のセグメントラティス

さは，同一セグメント内で，才1位の尤度値との差がある閾値以内の音韻を高々5位まで許している．

なお，音韻の種類は，入力に表われるすべての音韻を用いており，5個の母音と18個の子音の合計23種類である．すなわち，

母音…/a，i，u，e，o/

子音…/m，n，b，d，g，z，zy，r，w，y，
p，t，k，ky，s，h，hy，N/

また，VCV音節の標準パターンは，これらの音韻の組合せでできるすべてのものを用意しており，V型が6個（撥音/N/は母音と同様に扱っている），VV型が36個，VCV型が420個の合計462種類である．

2.2 言語処理(6)(7)

言語処理部のブロック図を図6に示す．入力の音韻ラティスは，まず，どの予約項目の節であってどのような単語で構成されているか予測される．この予測をす

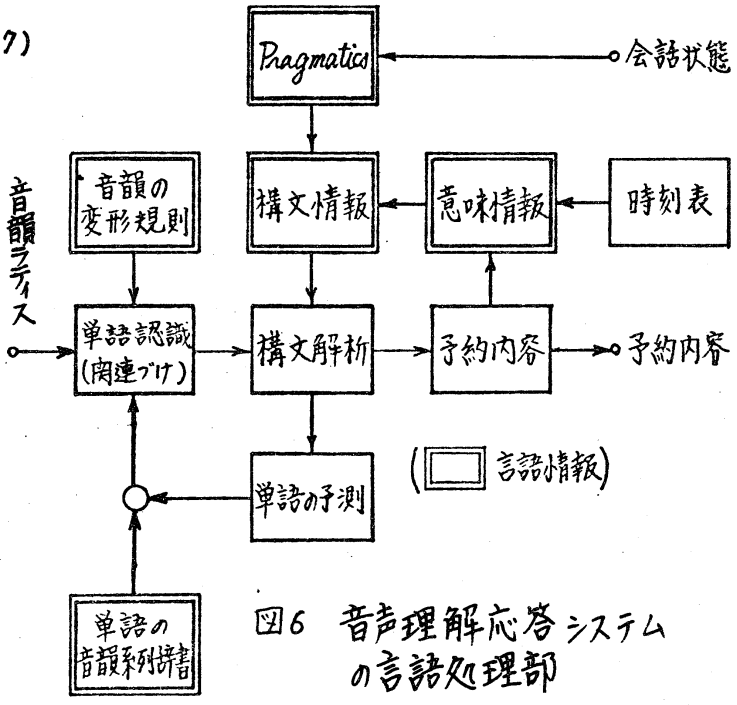


図6 音声理解応答システムの言語処理部

るための構文解析には，座席予約の会話に関する構文，意味 pragmatics などの情報を用いる．こうして予測された単語について，音韻変形規則を用いて，音韻ラティスと単語辞書との関連づけによる単語認識を行なう．その結果は，次の単語を予測するための構文解析に用いられる．こうして得られた入力文の予約内容が音声応答部へ送られる．

現在，音韻変形規則，単語辞書，構文情報を利用した単語認識と構文解析について検討しているので，それらについて次に述べる．

（1）音韻変形規則

音韻ラティスは，図5にみられるように，発声器官の習性による音韻の脱落や調音結合，音響処理部で生じた種々の誤りなどを含む．このような音韻ラティスに対して，表2に示される7種類のタイプの音韻変形規則を適用する．変形規則は，表2にみられるように，文脈依存型とし，各変形の生じ

表2 音韻変形規則

変形規則の種類	変形規則の型
1. 音韻の類似性	$xv \rightarrow yv \quad (p)$
2. 調音結合	$uxv \rightarrow uyv \quad (p)$
3. 無声化母音の挿入	$x \rightarrow y_1 y_2 y_3 \quad (p)$
4. 文頭の脱落子音の挿入	$\#\# \rightarrow \#\# y \quad (p)$
5. 音韻の挿入	$uv \rightarrow uyv \quad (p)$
6. 音韻の読みとばし	$uxv \rightarrow uv \quad (p)$
7. セグメントミスの訂正	$x_1 x_2 x_3 \rightarrow y \quad (p)$

- x, y, u, v：音韻記号（u, vは空記号もとりうる）．
- ##：文頭を表わす記号．
- p：変形規則を適用したときのペナルティ．

やすさに応じてその規則を適用したときにペナルティを課する．現在，156種類の変形規則を用意している．

なお，「4時，9時，1分，10分，1人，2人，・・・」などにおけるように，数字と数助詞の結合によって音便変化を生じる単語に対しては，複数個の音韻系列辞書を用意することとし，そのための音韻変形規則は作らない方が良いと考えている．

（2）単語辞書

単語辞書は，たとえば，名古屋（NAGOYA）のように，通常の音韻系列で表現されている．このような形式にすることによって，辞書の登録が極めて容易であり，しかも，発声者に全く依存しないという利点をもっている．辞書に登録する単語は，名詞38，助詞15，副詞2の合計55単語であり，予約項目に関係のない言葉は登録しないでリジェクトする．

（3）単語認識（関連づけ）

音韻ラティスは音韻系列の集合である．問題は，この音韻系列に音韻変形規則を適用して順次変形していき，それが予測された単語の音韻系列と完全に一致するかどうかを調べることである．このことを音韻系列の関連づけとよぶ．

具体的には，音韻ラティス中で単語の始まりのセグメント

が指定されると，そこから適用可能な音韻変形規則を，ペナルティの総和が一定値以内という条件のもとで順次適用し，関連づけが可能な単語の終りのセグメントをすべて求める．この関連づけを，3．で述べる理由により，*tree search* で行なう．*tree search* には一般に *breadth-first method* と *depth-first method* があるが，これも3．で述べる理由により，ここでは後者を用いることとする．

tree search を能率良く行なうために，*push-down stack* とメモリを用いる．*push-down stack* には，関連づけの途中の経過を残すために，*tree* の先に進むときには次の8組のパラメータを *push-down* する．

（S1）音韻ラティスのセグメントの番地
（S2）〃　セグメンテーションの番地
（S3）〃　音韻の候補の番地
（S4）単語辞書の音韻の番地
（S5）音韻変形規則の種類
（S6）〃　番地
（S7）ペナルティの総和
（S8）認識経過を示すリストの番地

また，メモリには，関連づけの途中で適用する変形規則がなかったりペナルティの総和が一定値を越えた場合，その都度

次の3組のパラメータを記憶する.

（M1）音韻ラティスのセグメントの番地

（M2）単語辞書の音韻の番地

（M3）ペナルティの総和

これらの情報を用いて，関連づけにおいて同じことを繰返さないようにする．すなわち，関連づけの途中で前述の8組のパラメータを *push-down* して *tree* の先に進むときは，現在の状態をメモリの内容と比較して，もし，

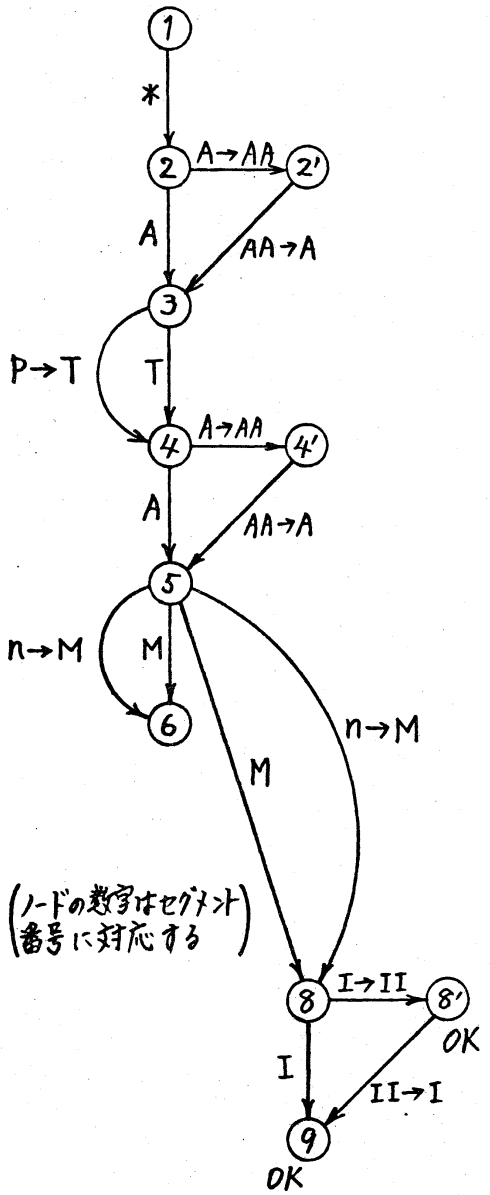
（S1）＝（M1）

（S4）＝（M2）

（S7）≧（M3）

ならば，先に進んでも無駄であるので，*push-down stack* を *pop-up* して関連づけをやり直す．こうすることによって，関連づけのステップ数を大中に削減することができる．

図7は，このような関連づけにおける *tree search* の様子と，*push-down stack* の動きおよびメモリ利用の様子の例である．この例は，図5の音韻ラティスにおいて#1セグメントを単語の始まりとして，駅名の単語〝熱海（ATAMI）〟を関連づけた場合を示している．



(a) tree search の様子

ノード番号	音韻の 読み込み	音韻の 変形	スタックの動き プッシュダウン	関連づけ	途中の記憶結果	記憶の利用
1	*		↙			
2	A		↙			
3	T		↙			
4	A		↙			
5	M		↙			
6			↗		M_1	
5	M		↙			
8	I		↙			
9			↗	OK	M_2	
8		I→II	↙			
8'	I	II→I	↙	OK		
9			↗			M_2
8'			↗		M_3	
8			↗		M_4	
5	n	n→M	↙			
6			↗			M_1
5	n	n→M	↙			
8			↗			M_4
5			↗		M_5	
4		A→AA	↙			
4'	A	AA→A	↙			
5			↗			M_5
4'			↗		M_6	
4			↗		M_7	
3	P	P→T	↙			
4			↗			M_7
3			↗		M_8	
2	.	A→AA	↙			
2'	A	AA→A	↙			
3			↗			M_8
2'			↗		M_9	
2			↗		M_{10}	
1			↗		M_{11}	

(b) プッシュダウンスタックの動きとメモリ利用の様子

図7 関連づけの例

(図5の音韻テディスにおいて#1セグメントを単語の始まりとして，駅名"熱海(ATAMI)"を関連づけた場合)

(4) 構文情報

予約項目の節ごとに構文情報を記述する．構文を表現するために，次の5つの演算子を用いる．この演算子は，単語認識を試みる順序も与えている．

AND：(AND x_1 x_2 ・・・ x_n)

$x_1, x_2, \cdots, x_n$ のすべてがこの順序で存在する．x_1 から順次 x_n まで試みること．

OR　：(OR x_1 x_2 ・・・ x_n)

$x_1, x_2, \cdots, x_n$ のいずれかが存在する．x_1 から順次成功するまで試みること．

OPT：(OPT x)

x が存在するかもしれない．まず，x が存在するとして試みること．

＊　：(＊ X)

Xがリストの名前であることを示す．リストXを代入すること．

SEM：　SEM(X a_1 a_2 ・・・ a_n)

Xというセマンティクスを挿入する．セマンティクルーチンXを動作させること．a_i はパラメータである．

こうして表現された発駅と着駅に関する節の構文を図8に示す．

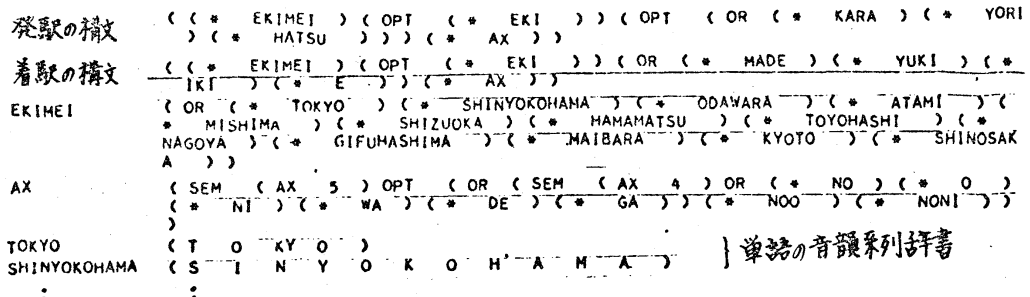


図8 構文情報の表現例（発駅と着駅）

(5) 構文解析

ここでの構文解析の目的は単語を予測することである．したがって，構文解析は top - down 形式で行なう．

図8のように表現された構文情報を depth - first でたどり，単語の音韻系列に到達すると，その単語が入力されたと予測して関連づけを実行する．関連づけの結果が成功であるか失敗であるかに応じて次のステップに進む．この場合にも，関連づけの場合と同様に，構文解析の途中の結果を push - down stack とメモリに記憶して，その後の構文解析の過程で利用することによって，同じことを繰返さないようにしている．

2.3 音声応答

ここでは，会話状態遷移図に従って，予約内容に応じた応答文を生成し，その応答文の音声を合成することが必要である．詳細は省略する．

3. 音声による質問応答システムにおける時系列パターン

音声による質問応答システムを時系列パターン認識システ

ムとしてみると，図9のようになり，次の3種類の時系列パターンを扱っている．

- VCV音声の認識における特徴パラメータ系列パターン．

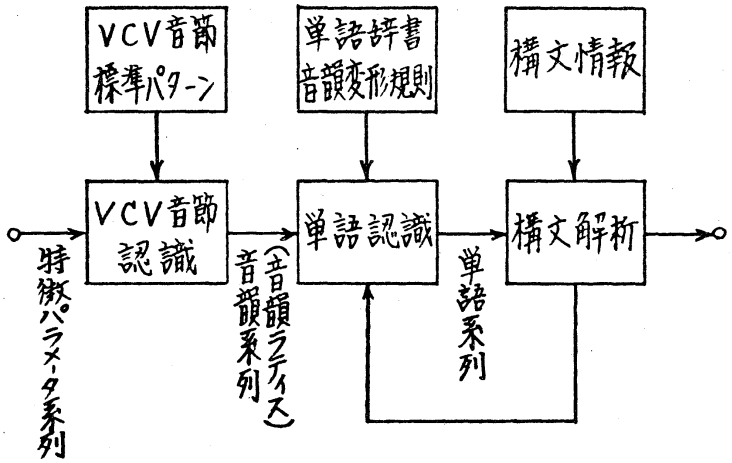


図9 時系列パターン認識システムとしてみた音声による質問応答システム

- 単語の認識における音韻系列（音韻ラティス）パターン．
- 構文解析（節の認識）における単語系列パターン

時系列パターンの認識を行なうとき，次のようないくつかの選択すべき事柄があり，各時系列パターンの性質にあった処理方法を選ぶことが大切である．

(1) マッチングの演算をDPで行なうか tree search で行なうか．

(2) tree search を breadth-first で行なうか depth-first で行なうか．

(3) マッチングの順序を left-right で行なうか anchor point search で行なうか．

(4) マッチングの手順を bottom-up で行なうか top-down で行なうか．

(1) に関しては，時系列パターンの変形の性質が一様である場合にはDPが適し，多様である場合には tree search が適すると考えている．そして，tree search で行なうときには，DPにおけると同様に，途中の結果をその後の過程で有効に利用することが大切である．

VCV音節の認識における特徴パラメータ系列パターンは主として発声速度による変形を受け，この変形は比較的一様であるのでDPが適している．それに対して，単語の認識における音韻系列パターンは，入力が音韻のあいまいさとセグメンテーションのあいまいさの両方を許した音韻ラティスであること，音韻変形規則がいろいろの種類の変形規則からなり，しかも，文脈依存型であること，などのために極めて多様な変形を行なう必要があるので tree search が適すると考えている．構文解析における単語系列パターンについても，同様の理由により，tree search で行なっている．

(2) に関しては，depth-first をとっている．その理由は，記憶容量が少なくてすむことと，depth-first でもそれを続けることによって breadth-first 的な search を行なうこともできるからである．

(3) に関しては left-right で行ない，(4) に関しては構文解析のときのみ top-down で行なっている．

表3　時系列パターンの処理方法の比較

	特徴パラメータ系列	音韻系列	単語系列
認識対象	VCV音節	単語	節
マッチングの演算	DP	tree search	tree search
tree searchの方法	(breadth-first的)	depth-first	depth-first
マッチングの順序	left-right	left-right	left-right
マッチングの手順	bottom-up	bottom-up	top-down

ここでのシステムにおける各時系列パターンに対して用いている処理方法をまとめて表3に示す.

4. むすび

現在検討を進めている音声による質問応答システムの概要に関して，入力の音響処理部と言語処理部について述べた．そして，このシステムを時系列パターン認識システムとしてみると，特徴パラメータレベル，音韻レベル，単語レベルの3種類の時系列パターンを扱っていることを述べ，各時系列パターンの処理方法を比較検討した．

ここで述べた処理方法は現在行なっている方法であって，今後，座席予約の会話に関する意味やpragmaticsなどの情報もとり入れるとともに，各処理方法についてもさらに検討を加える予定である．

(謝辞) 御指導賜わる野田基礎研究部長，斉藤オ々研究室長，および，熱心に討論していただくオ々研究室の諸氏に深謝する．

(文献)

(1) IEEE Symposium on Speech Recognition (1974-04).

(2) 電四連大パネル討論，音声認識における言語情報の利用 (1974-10).

(3) 好田：言語情報を利用した文章音声認識の検討，昭49電四連大199 (1974-10).

(4) 中津，好田：連続音声のセグメント化と音韻系列への変換，音声研究会資料S74-25 (1974-12).

(5) 中津，好田：VCV音節を単位とした単語音声の認識，信学会研究会資料PRL73-63 (1973-09).

(6) 鹿野，好田：構文情報を利用した文章音声認識 (音声による質問応答システム)，音学講論2-2-21 (1974-10).

(7) 鹿野，好田：構文情報を用いた会話音声の認識，情報処理学会オ15回大会38 (1974-12).